J. Inverse Ill-Posed Probl. **19** (2011), 859–879
DOI 10.1515/JIIP.2011.052

Enhancing linear regularization to treat large noise

Peter Mathé and Ulrich Tautenhahn

Abstract. For solving linear ill-posed problems with noisy data, regularization methods are required. In this paper we study regularization under general noise assumptions containing large noise and small noise as special cases. We derive order optimal error bounds for an extended Tikhonov regularization by using some pre-smoothing. This accompanies recent results by the same authors, *Regularization under general noise assumptions*, Inverse Problems 27:3, 035016, 2011.

Keywords. Ill-posed problems, inverse problems, regularization, Hilbert scales, order optimal error bounds, large noise, small noise.

2010 Mathematics Subject Classification. 65J20, 47A52.

1 Introduction

Ill-posed problems arise in several context and have important applications in science and engineering (see e.g. [1, 5, 10]). In this paper we consider ill-posed problems

$$Ax = y \tag{1.1}$$

with bounded linear operators $A: X \to Y$ mapping between infinite dimensional Hilbert spaces X and Y with inner products $(\cdot, \cdot)$ and norms $\|\cdot\|$. To simplify the presentation, we assume that the operator A is injective. The extension to the more general case where A is not injective can be handled analogously to [5]. We further assume that the range $\mathcal{R}(A)$ is non-closed and that y belongs to $\mathcal{R}(A)$ such that (1.1) has a unique solution $x^\dagger \in X$. We have to add assumptions made on the noisy data y^δ, and we assume that there is a real number q for which

$$\|(AA^*)^{q/2}(y - y^\delta)\| \le \delta, \tag{1.2}$$

with known noise level δ. We call the case $q > 0$ *large noise case*, and we shall call the case $q < 0$ *small noise case*. Traditionally, regularization methods are studied under the noise condition (1.2) with $q = 0$.

We further assume that the unknown solution $x^\dagger$ fulfills a polynomial source condition, i.e., with some (generally unknown) $p > 0$,

$$x^\dagger \in M_{p,E} := \left\{ x \in X \;\middle|\; x = (A^*A)^{p/2} v,\ \|v\| \le E \right\}. \tag{1.3}$$

Remark 1.1. Both of the above assumptions can be generalized by replacing the power type condition, as e.g. $x = (A^*A)^{p/2}v$, by some general continuous non-decreasing function $\varphi: (0, \|A^*A\|] \to (0, \infty)$, hence that $x = \varphi(A^*A)v$, and we shall indicate this in Section 3. However, the main picture can be seen under these power-type conditions, and we concentrate on such.

Such smoothness conditions correspond to *Hilbert scales*, indexed by a parameter $r \in \mathbb{R}$, where the norm $\|\cdot\|_r$ is given by

$$\|x\|_r := \| \left(A^*A\right)^{-r/2} x\|, \quad \text{provided that } x \in \mathcal{R}(\left(A^*A\right)^{r/2}).$$

Under above assumption on the noise and the smoothness we can find the best possible accuracy of reconstruction as this is captured in the notion of the modulus of continuity. Let $R : Y \to X$ be an arbitrary method. Then, the quantity

$$\Delta(\delta, R) = \sup\left\{\|R(y^\delta) - x^\dagger\| \;\middle|\; \|(AA^*)^{q/2}(y - y^\delta)\| \le \delta,\ x^\dagger \in M_{p,E}\right\}$$

is called the *worst case error* of the method R on the set $M_{p,E}$ under the noise condition (1.2). An optimal method R_{opt} is characterized by

$$\Delta(\delta, R_{\mathrm{opt}}) = \inf_R \Delta(\delta, R),$$

and this quantity is called *best possible worst case error* on the set $M_{p,E}$ under the noise condition (1.2). It was shown in Theorem 4 in [8] that

$$\inf_R \Delta(\delta, R) \le E^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}} \quad \text{for } (p, q) \in (0, \infty) \times (-1, \infty), \tag{1.4}$$

and that equality holds only if δ/E is an element of the spectrum $\sigma(G)$ of the operator $G = (A^*A)^{(p+q+1)/2}$. Therefore, the right hand side of (1.4) is the benchmark for the best possible accuracy for identifying $x^\dagger$ from noisy data under the noise condition (1.2) and the smoothness condition (1.3).

Remark 1.2. Notice that necessarily we require that $q > -1$, since for $q \le -1$ the problem is well-posed.

The objective of this study can be comprised in the following questions: First, can we achieve this (optimal) order of reconstruction by means of linear regularization? The following was observed on a previous study [4]. If we use Tikhonov regularization, when the approximate solutions is given as

$$x_\alpha^\delta = \left(A^*A + \alpha I\right)^{-1} A^* y^\delta, \tag{1.5}$$

under large noise, when $\|A^*(y - y^\delta)\| \le \delta$, then the best possible order of reconstruction can be achieved only for solution smoothness $x^\dagger \in M_{1,E}$, in contrast to the usual situation when optimal reconstruction is possible up to $x^\dagger \in M_{2,E}$.

The second question is about *a posteriori parameter choice*. Do a posteriori rules work, or can such be modified to work under more general noise assumptions? Again, [4] answers this question in the affirmative for the Lepskiĭ (balancing) principle. Here we shall consider other more traditional parameter choice rules, as the discrepancy principle and the Raus–Gfrerer rules.

Within this study we will show that the following modification of Tikhonov regularization will be useful under general noise assumptions. We start from the symmetrized noisy equation

$$A^* y^\delta = A^* A x + \delta A^* \xi, \tag{1.6}$$

and we apply some power of $A^* A$ to both sides. Precisely, for some $s > -1$ we let

$$(A^* A)^s A^* y^\delta = (A^* A)^{s+1} x + \delta (A^* A)^s A^* \xi. \tag{1.7}$$

If we now apply Tikhonov regularization, based on the operator $(A^* A)^{s+1}$ then we consider as approximate solution the family

$$x_\alpha^\delta = \left((A^* A)^{s+1} + \alpha I\right)^{-1} (A^* A)^s A^* y^\delta, \quad \alpha > 0. \tag{1.8}$$

In (1.5), (1.8), the parameter α is the regularization parameter, and s is some properly chosen number. From the viewpoint of computational amount, s is generally an integer in case $A \neq A^*$, and $2s$ is generally an integer in case $A = A^*$.

Remark 1.3. The procedure stated above may be considered as *pre-smoothing* of the raw data y^δ, in order to place them in the target space Y. Then we apply standard Tikhonov regularization (1.5) to this enhanced equation. We stress that $s = 0$ corresponds to the ordinary Tikhonov scheme. For $s = -1/2$ the operator

$$Q := (A^* A)^{-1/2} A^* \colon Y \to X$$

is an isometry, and the equation (1.7) reduces to $Q y^\delta = |A|\, x + \delta Q \xi$, which for non-negative self-adjoint $A = A^* > 0$ gives the original equation. In this case Tikhonov regularization reduces to $x_\alpha^\delta = (A + \alpha I)^{-1} y^\delta$, $\alpha > 0$, which corresponds to Lavrent′ev regularization of non-negative equations.

As already mentioned, convergence rate results under large noise have been obtained before. In [9] Tikhonov regularization for the special case $q = 1$ in (1.2), $p = 1$ or $p = 2$ in (1.3) and $s = 0$ in (1.8) has been treated under *a priori* parameter choice for α. Generalizations of the results from [9] may be found in the papers [3, 4, 8].

Here we extend these studies by using general linear regularization as introduced in equation (1.8), and generalizations thereof. In particular, we are interested in (p, q)-ranges that, under the noise condition (1.2) and the smoothness condition (1.3), guarantee order optimal error bounds $\|x_\alpha^\delta - x^\dagger\| = O(\delta^{p/(p+q+1)})$. We shall prove in Theorem 2.2 below that, given p, q, order optimal error bounds can be guaranteed by choosing α according to the *a priori* parameter choice for $\alpha \asymp \delta^{(2s+2)/(p+q+1)}$ whenever

$$s \geq \max\left\{\frac{p-2}{2}, \frac{q-1}{2}\right\}. \tag{1.9}$$

Thus, the choice of s has two-fold implications. Tikhonov regularization is capable to react to larger noise (larger value for q) and to higher smoothness (larger value of p), beyond the usual limitation $p \leq 2$ for Tikhonov regularization. Thus, by enlarging $s > -1$ we increase the qualification of the regularization described in (1.5) by pre-smoothing the data.

In case of the *a posteriori* parameter choice by the discrepancy principle we shall prove order optimal reconstruction for

$$s \geq \max\left\{\frac{p+q-1}{2}, \frac{q-1}{2}\right\} \geq \max\left\{\frac{p-2}{2}, \frac{q-1}{2}\right\},$$

since necessarily we have that $q \geq -1$, see Remark 1.2.

In case of choosing α *a posteriori* by either the balancing principle or the Raus–Gfrerer rule, the choice of s as in (1.9) suffices to guarantee order optimality, and this is indicated for the balancing principle in Section 2.6, and it is established in Theorem 2.7 for the Raus–Gfrerer rule. We conclude this study with a discussion on generalizations, in particular in view of other recent publications in this direction.

2 The error of Tikhonov-type regularization

Recall that we use the modification of the Tikhonov regularization as introduced in (1.8). Our analysis adapts standard methods from general regularization theory, where the special case $q = 0$ is treated, see e.g. [5, 11].

2.1 Error decomposition

As usual we decompose the error into the regularization error and the term for noise propagation, as

$$\|x^\dagger - x_\alpha^\delta\| \leq \|x^\dagger - x_\alpha\| + \|x_\alpha - x_\alpha^\delta\|, \quad \alpha > 0, \tag{2.1}$$

where, for $\alpha > 0$,

$$x_\alpha := \left((A^*A)^{s+1} + \alpha I\right)^{-1} (A^*A)^s A^* y = \left((A^*A)^{s+1} + \alpha I\right)^{-1} (A^*A)^{s+1} x^\dagger.$$

Let us introduce the regularization function g_α and the residual function r_α by

$$g_\alpha(\lambda) = \frac{1}{\lambda + \alpha} \quad \text{and} \quad r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda) = \frac{\alpha}{\lambda + \alpha}.$$

By using (1.3), we have for the regularization error (bias) the estimate

$$\begin{aligned} \|x^\dagger - x_\alpha\| &= \|r_\alpha((A^*A)^{s+1})(A^*A)^{p/2}v\| \\ &\le E \sup_{\lambda>0} \frac{\alpha\lambda^{p/2}}{\lambda^{s+1} + \alpha} \\ &\le E\alpha^{\frac{p}{2s+2}} \qquad \text{for } 0 < p \le 2s+2. \end{aligned} \tag{2.2}$$

By using (1.2), we have for the noise propagation error the estimate

$$\begin{aligned} \|x_\alpha - x_\alpha^\delta\| &= \|g_\alpha((A^*A)^{s+1})(A^*A)^s A^*(y - y^\delta)\| \\ &\le \delta \sup_{\lambda>0} \frac{\lambda^{s+1/2-q/2}}{\lambda^{s+1} + \alpha} \\ &\le \delta\alpha^{-\frac{q+1}{2s+2}} \qquad \text{for } -1 < q \le 2s+1. \end{aligned} \tag{2.3}$$

For both estimates (2.2) and (2.3) we used the substitution $\lambda^{s+1} = \alpha t$ and the fact that

$$\sup_{t>0} \frac{t^\nu}{t+1} \le 1 \quad \text{for } 0 \le \nu \le 1.$$

We summarize this as follows.

Proposition 2.1. *Suppose that $x^\dagger \in M_{p,E}$ for some $p > 0$ and that the noise obeys* (1.2) *for some $q \ge -1$. If $s \ge \max\left(\frac{p-2}{2}, \frac{q-1}{2}\right)$, then*

$$\|x^\dagger - x_\alpha^\delta\| \le E\alpha^{\frac{p}{2s+2}} + \delta\alpha^{-\frac{q+1}{2s+2}}, \quad \alpha > 0.$$

2.2 A priori parameter choice

The error bound in Proposition 2.1 allows for an order optimal choice of the regularization parameter α by equating both summands.

Theorem 2.2. *Assume the noise condition* (1.2)*, the smoothness condition* (1.3)*, let x_α^δ be defined by* (1.8)*, and let α be chosen as $\alpha = \left(\frac{\delta}{E}\right)^{(2s+2)/(p+q+1)}$. Then,*

$$\|x_\alpha^\delta - x^\dagger\| \le 2E^{\frac{q+1}{p+q+1}}\delta^{\frac{p}{p+q+1}}, \tag{2.4}$$

provided that $s \ge \max\left(\frac{p-2}{2}, \frac{q-1}{2}\right)$.

We discuss the above result. First notice that the same rate $O\left(\delta^{\frac{p}{p+q+1}}\right)$ is obtained regardless of the chosen parameter s, provided that it is large enough. However, the optimal parameter does depend on the value of s, and it is larger the smaller s is. Since large regularization parameters correspond to better stability, it is desirable to choose s as small as possible. However, this seriously restricts the values p in the smoothness classes $M_{p,E}$.

In the special case $s = 0$ and $q = 0$ we obtain the well-known result that the best possible convergence rate for ordinary Tikhonov regularization is of order $O(\delta^{2/3})$. In the special case $s = -1/2$ and $q = 0$ we obtain the well-known result that the best possible convergence rate for Lavrent′ev regularization is of order $O(\delta^{1/2})$.

2.3 Discrepancy principle

If the constants p and E in the *a priori* parameter choice of Theorem 2.2 are unknown, then *a posteriori* rules for choosing α should be used. Here we propose a modification of the classical discrepancy principle. Due to the noise condition (1.2) it makes sense to choose α as the solution of the nonlinear equation

$$d(\alpha) := \left\| (AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta) \right\| = C\delta, \tag{2.5}$$

with some constant $C \geq 1$. However, if q is not an integer, then the numerical realization of this principle is expensive. The function d may be rewritten as

$$d(\alpha) = \left\| \alpha \left((AA^*)^{s+1} + \alpha I \right)^{-1} (AA^*)^{q/2} y^\delta \right\|.$$

From this representation we conclude that d is monotonically increasing and obeys

$$\lim_{\alpha \to 0} d(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha \to \infty} d(\alpha) = \|(AA^*)^{q/2} y^\delta\|.$$

From these properties we conclude that equation (2.5) has a unique positive solution α_D provided $\|(AA^*)^{q/2} y^\delta\| > C\delta$.

Remark 2.3. It is well known that the zero element $x_\alpha^\delta := 0$ is an order optimal solution if the data are small, i.e.,

$$\|(AA^*)^{q/2} y^\delta\| \leq C\delta,$$

see e.g. [2, Lemma 4.6].

Otherwise, the choice of α is important. Given any value of s that obeys (1.9), and any value $q \geq -1$, when does α_D yield order optimal convergence?

Theorem 2.4. *Assume the noise condition* (1.2)*, the smoothness condition* (1.3)*, let x_α^δ be defined by* (1.8) *and let $\alpha = \alpha_D$ be chosen by* (2.5) *with $C > 1$. Then,*

$$\|x_\alpha^\delta - x^\dagger\| \le \left((C+1)^{\frac{p}{p+q+1}} + (C-1)^{-\frac{q+1}{p+q+1}}\right) E^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}, \tag{2.6}$$

provided that $0 < p \le 2s + 1 - q$.

Proof. From (2.5), (1.2), (1.3) and $|r_\alpha(\lambda)| \le 1$ we have

$$\begin{aligned} C\delta &\le \left\|(AA^*)^{q/2} r_\alpha((AA^*)^{s+1})(y - y^\delta)\right\| + \left\|(AA^*)^{q/2} r_\alpha((AA^*)^{s+1}) y\right\| \\ &\le \delta + E \sup_\lambda \frac{\alpha \lambda^{(p+q+1)/2}}{\lambda^{s+1} + \alpha} \\ &\le \delta + E\alpha^{\frac{p+q+1}{2s+2}} \qquad \text{for } -1 < p + q \le 2s + 1. \end{aligned} \tag{2.7}$$

From (2.3) and (2.7) we obtain that for $\alpha = \alpha_D$ the noise propagation error can be estimated by

$$\|x_\alpha^\delta - x_\alpha\| \le \left(\frac{E}{C-1}\right)^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}. \tag{2.8}$$

For estimating the regularization error (bias) for $\alpha = \alpha_D$ we proceed in different steps. In the first step we use (2.5), (1.2) and $|r_\alpha(\lambda)| \le 1$ and obtain

$$\begin{aligned} \|x^\dagger - x_\alpha\|_{-q-1} &= \|(AA^*)^{q/2} r_\alpha((AA^*)^{s+1}) y\| \\ &\le \|(AA^*)^{q/2} r_\alpha((AA^*)^{s+1}) y^\delta\| \\ &\qquad + \|(AA^*)^{q/2} r_\alpha((AA^*)^{s+1})(y - y^\delta)\| \\ &\le (C+1)\delta. \end{aligned} \tag{2.9}$$

In the second step we use (1.3) and $|r_\alpha(\lambda)| \le 1$ and obtain

$$\|x^\dagger - x_\alpha\|_p = \|r_\alpha((A^*A)^{s+1}) v\| \le E. \tag{2.10}$$

In the third step we use (2.9) and (2.10), apply the interpolation inequality

$$\|x\|_r \le \|x\|_{-a}^{(b-r)/(b+a)} \|x\|_b^{(a+r)/(b+a)} \tag{2.11}$$

that holds true for $r \in [-a, b]$, $a + b \ne 0$, and obtain

$$\|x^\dagger - x_\alpha\| \le E^{\frac{q+1}{p+q+1}} [(C+1)\delta]^{\frac{p}{p+q+1}} \qquad \text{for } -p < 0 \le q + 1. \tag{2.12}$$

Now, estimate (2.6) follows from (2.8) and (2.12). □

We already noticed at the end of Section 1 that, given $s > -1$, the range for values of p for which order optimal reconstruction is obtained by using the discrepancy principle is smaller than the range under a priori choice of the parameter.

In the special case $s = 0$ and $q = 0$ we obtain the well-known result that the best possible convergence rate for ordinary Tikhonov regularization with α chosen by the discrepancy principle is of order $O(\delta^{1/2})$. In the special case $s = -1/2$ and $q = 0$ we obtain the well-known result that there does not exist a convergence rate for Lavrent′ev regularization with α chosen by the discrepancy principle.

2.4 Discrepancy principle revisited

In this subsection we show that for some restricted (p, q)-range the estimate (2.6) can be improved. In particular, we will derive some sharper estimate which is also valid in the case $C = 1$. We start our study with some auxiliary result.

Proposition 2.5. *Assume the noise condition* (1.2)*, the smoothness condition* (1.3)*, let x_α^δ be defined by* (1.8) *and let α be chosen by* (2.5) *with $C \geq 1$. Then,*

$$\|x_\alpha^\delta - x^\dagger\|_{s-q}^2 \leq E\|x_\alpha^\delta - x^\dagger\|_{2s-2q-p}. \tag{2.13}$$

Proof. We use the abbreviation $z_\alpha^\delta = x^\dagger - x_\alpha^\delta$, use in addition the representation $\alpha x_\alpha^\delta = (A^*A)^s A^*(y^\delta - Ax_\alpha^\delta)$ that follows from (1.8) and obtain

$$\begin{aligned}
\|z_\alpha^\delta\|_{s-q}^2 &= \left(x^\dagger, (A^*A)^{q-s} z_\alpha^\delta\right) - \left(x_\alpha^\delta, (A^*A)^{q-s} z_\alpha^\delta\right) \\
&= \left(x^\dagger, (A^*A)^{q-s} z_\alpha^\delta\right) \\
&\qquad + \frac{1}{\alpha}\left((AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta), (A^*A)^{q/2}(Ax^\dagger - Ax_\alpha^\delta)\right) \\
&= \left(x^\dagger, (A^*A)^{q-s} z_\alpha^\delta\right) \\
&\qquad + \frac{1}{2\alpha}\Big(\|(AA^*)^{q/2}(y - y^\delta)\|^2 - \|(AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta)\|^2 \\
&\qquad\qquad - \|(AA^*)^{q/2}(y - Ax_\alpha^\delta)\|^2\Big).
\end{aligned} \tag{2.14}$$

Due to (1.2) and (2.5), the expression in the brackets is negative. Hence, by using (1.3) we have

$$\|z_\alpha^\delta\|_{s-q}^2 \leq \left((A^*A)^{p/2}v, (A^*A)^{q-s} z_\alpha^\delta\right).$$

From this estimate we obtain (2.13). □

Theorem 2.6. *Assume the noise condition* (1.2)*, the smoothness condition* (1.3)*, let* x_α^δ *be defined by* (1.8) *and let* α *be chosen by* (2.5) *with* $C \geq 1$*. If* $s \geq q$*, then*

$$\|x_\alpha^\delta - x^\dagger\| \leq E^{\frac{q+1}{p+q+1}} \left[(C+1)\delta\right]^{\frac{p}{p+q+1}}, \tag{2.15}$$

provided that $0 \leq s - q \leq p \leq 2s + 1 - q$.

Proof. We use the abbreviation $z_\alpha^\delta = x^\dagger - x_\alpha^\delta$ and obtain due to (2.5) and (1.2) the estimate

$$\begin{aligned} \|z_\alpha^\delta\|_{-q-1} &= \|(AA^*)^{q/2}(Ax_\alpha^\delta - y)\| \\ &\leq \|(AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta)\| + \|(AA^*)^{q/2}(y - y^\delta)\| \\ &\leq (C+1)\delta. \end{aligned} \tag{2.16}$$

Next, we derive an inequality that relates the three norms $\|z_\alpha^\delta\|_{-q-1}$, $\|z_\alpha^\delta\|_{s-q}$ and $\|z_\alpha^\delta\|_{2s-2q-p}$. Using the interpolation inequality (2.11) with $r = 2s - 2q - p$, $a = q + 1$ and $b = s - q$ yields

$$\|z_\alpha^\delta\|_{2s-2q-p} \leq \|z_\alpha^\delta\|_{-q-1}^{(p+q-s)/(s+1)} \|z_\alpha^\delta\|_{s-q}^{(2s+1-p-q)/(s+1)}. \tag{2.17}$$

Manipulation of the three estimates (2.13), (2.16) and (2.17) yields

$$\|z_\alpha^\delta\|_{s-q} \leq E^{\frac{s+1}{p+q+1}} \left[(C+1)\delta\right]^{\frac{p+q-s}{p+q+1}}. \tag{2.18}$$

Finally, applying again the interpolation inequality (2.11) with $r = 0$, $a = q + 1$ and $b = s - q$ yields together with (2.16) and (2.18) the estimate (2.15). □

Notice, that the same maximal smoothness is required to retain order optimality. However, we may apply the discrepancy principle with $C = 1$ only, if the smoothness also is bounded from below. In the classical case, when $s = q = 0$ then this cannot be seen. However, if $q = 0$ but $s > 0$, then this gives a lower bound for the (implicitly) assumed smoothness.

2.5 Raus–Gfrerer rule

The Raus–Gfrerer parameter choice rule overcomes the early saturation of the discrepancy principle. In the traditional setup, for Tikhonov regularization

$$g_\alpha(\lambda) = 1/(\lambda + \alpha),$$

with residual function (bias) given as $r_\alpha(\lambda) = \alpha/(\lambda + \alpha)$, and when $s = 0$, $q = 0$, the Raus–Gfrerer rule proposes to choose the regularization parameter α as the solution of the nonlinear equation

$$d_{\mathrm{RG}}(\alpha) := \left\| r_\alpha(AA^*)^{1/2}(Ax_\alpha^\delta - y^\delta)\right\| = C\delta \tag{2.19}$$

and some constant $C > 1$ leading to regularized solutions that are order optimal for the maximal range $p \in (0,2]$ of solution smoothness. We are interested to generalize this rule for noise situations (1.2) with $q \neq 0$ and regularization methods (1.8) with $s \neq 0$ in such a way that the regularized solutions are order optimal for the maximal (p,q)-range $(p,q) \in (0, 2s+2] \times (-1, 2s+1]$ of Theorem 2.2. We assign two operators $G_s := (AA^*)^{s+1}$ and $H_s := (A^*A)^{s+1}$, and we generalize the Raus–Gfrerer rule as follows. Given some $\mu \geq 0$ we choose the regularization parameter α as the solution of the nonlinear equation

$$d_{\mathrm{RG}}(\alpha) := \left\| r_\alpha(G_s)^\mu(AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta)\right\| = C\delta. \tag{2.20}$$

Note that for $s = 0$ and $q = 0$, and $\mu = 1/2$, the rule (2.20) coincides with (2.19). The function d_{RG} may be rewritten in a more convenient form as

$$d_{\mathrm{RG}}(\alpha) = \left\| r_\alpha(G_s)^{1+\mu}(G_s)^{q/(2s+2)}y^\delta\right\|.$$

From this representation we conclude that d_{RG} is monotonically increasing and obeys the limit relations

$$\lim_{\alpha\to 0} d_{\mathrm{RG}}(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha\to\infty} d_{\mathrm{RG}}(\alpha) = \|(AA^*)^{q/2}y^\delta\|.$$

From these properties we conclude that equation (2.5) has a unique positive solution α_{RG} provided $\|(AA^*)^{q/2}y^\delta\| > C\delta$. For small data, $\|(AA^*)^{q/2}y^\delta\| \leq C\delta$, Remark 2.3 applies, and the zero solution yields order optimal rates.

In the general case, and for an appropriate choice of μ order optimality can be guaranteed.

Theorem 2.7. *Let x_α^δ be defined by* (1.8) *and $\alpha = \alpha_{\mathrm{RG}}$ be chosen by* (2.20) *with $C > 1$. Under the conditions* (1.2) *and* (1.3)*, let s cover the maximal range given by* (1.9)*. If $\mu := \frac{q+1}{2s+2}$, then*

$$\|x_{\alpha_{\mathrm{RG}}}^\delta - x^\dagger\| \leq \left((C+1)^{\frac{p}{p+q+1}} + (C-1)^{-\frac{q+1}{p+q+1}}\right) E^{\frac{q+1}{p+q+1}}\delta^{\frac{p}{p+q+1}}. \tag{2.21}$$

Proof. First notice that for s from (1.9) and with $\mu \geq \frac{q+1}{2s+2}$ we have that

$$p + q + 1 \leq (2s+2)(1+\mu).$$

Hence, we see from (2.20), (1.2), (1.3), and by $\|r_\alpha(G_s)\| \le 1$, that

$$\begin{aligned} C\delta &\le \left\| r_\alpha(G_s)^{1+\mu}(AA^*)^{q/2}(y-y^\delta) \right\| + \left\| r_\alpha(G_s)^{1+\mu}(AA^*)^{q/2} y \right\| \\ &\le \delta + E \sup_\lambda \left(\frac{\alpha}{\lambda^{s+1}+\alpha} \right)^{1+\mu} \lambda^{(p+q+1)/2} \\ &\le \delta + E\alpha^{\frac{p+q+1}{2s+2}} \qquad \text{since } p+q+1 \le (2s+2)(1+\mu). \end{aligned} \tag{2.22}$$

From (2.3) and (2.22) we obtain that for $\alpha = \alpha_{\mathrm{RG}}$ the noise propagation error can be estimated by

$$\|x_\alpha^\delta - x_\alpha\| \le \left(\frac{E}{C-1} \right)^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}. \tag{2.23}$$

For estimating the regularization error (bias) for $\alpha = \alpha_{\mathrm{RG}}$ we proceed in different steps. In the first step we use (2.20), (1.2) and $\|r_\alpha(G_s)\| \le 1$, and we obtain

$$\begin{aligned} \|r_\alpha^{1+\mu}(H_s)(A^*A)^{\frac{q+1}{2}} x^\dagger\| &= \|r_{\alpha_{\mathrm{RG}}}(G_s)^{1+\mu}(AA^*)^{q/2} A x^\dagger\| \\ &= \|r_{\alpha_{\mathrm{RG}}}(G_s)^{1+\mu}(AA^*)^{q/2} y\| \\ &\le \|r_{\alpha_{\mathrm{RG}}}(G_s)^{1+\mu}(AA^*)^{q/2} y^\delta\| \\ &\qquad + \|r_{\alpha_{\mathrm{RG}}}(G_s)^{1+\mu}(AA^*)^{q/2}(y-y^\delta)\| \\ &\le (C+1)\delta. \end{aligned} \tag{2.24}$$

In the second step we use (1.3), $p \le 2s+2$, and $\|r_\alpha(H_s)\| \le 1$ and obtain

$$\|r_\alpha^{1-p/(2s+2)}(H_s)v\| \le E. \tag{2.25}$$

In the third step we use (2.24) and (2.25), and we apply the moment inequality

$$\|B^a z\| \le \|B^b z\|^{a/b} \|z\|^{1-a/b} \quad (0 \le a \le b,\ b \ne 0) \tag{2.26}$$

with constants $a := p$, $b := p+q+1$, bounded operator

$$B := r_\alpha^{1/(2s+2)}(H_s)(A^*A)^{1/2}$$

and element $z := r_\alpha^{1-p/(2s+2)}(H_s)v$. If we take into account that due to (1.3) we have

$$\begin{aligned} B^a z &= r_\alpha^{\frac{p}{2s+2}}(H_s)(A^*A)^{\frac{p}{2}} r_\alpha^{1-\frac{p}{2s+2}}(H_s)v = r_\alpha(H_s)(A^*A)^{\frac{p}{2}} v \\ &= r_\alpha(H_s)x^\dagger = x^\dagger - x_\alpha, \\ B^b z &= r_\alpha^{\frac{p+q+1}{2s+2}}(H_s)(A^*A)^{\frac{p+q+1}{2}} r_\alpha^{1-\frac{p}{2s+2}}(H_s)v = r_\alpha^{1+\mu}(H_s)(A^*A)^{\frac{q+1}{2}} x^\dagger, \end{aligned}$$

then the moment inequality gives

$$\|x^\dagger - x_\alpha\| \le [(C+1)\delta]^{\frac{p}{p+q+1}} \, E^{\frac{q+1}{p+q+1}} \quad \text{for } 0 < p \le 2s+2. \tag{2.27}$$

Now, estimate (2.21) follows from (2.23) and (2.27). □

We discuss the above result in some detail. In the special case $s = 0$, $q = 0$ and $\mu = 1/2$ we obtain from Theorem 2.7 the well-known result that the best possible convergence rate for ordinary Tikhonov regularization with α chosen by the Raus–Gfrerer rule (2.19) is of order $O(\delta^{2/3})$.

In the special case $s = -1/2$ and $q = 0$ we obtain that the best possible rate for Lavrent′ev regularization with α chosen by rule (2.20) is of order $O(\delta^{1/2})$.

Let us finally compare both *a posteriori* rules (2.5) and (2.20). The fact that $\|r_\alpha(G_s)\| \le 1$ gives us that $d_{\mathrm{RG}}(\alpha) \le d(\alpha)$. Since both functions are monotonically increasing, we obtain that α_D and α_{RG} are related by $\alpha_D \le \alpha_{\mathrm{RG}}$. Furthermore, we know that for smooth solutions $x^\dagger$ that obey (1.3) with $p > 2s+1-q$ the convergence rate for α_{RG} is higher than the convergence rate for α_D. We conclude that the discrepancy principle (2.5) provides a regularization parameter α_D which is too small for too smooth solutions $x^\dagger$.

2.6 Balancing principle

Within the present context, if $q \ge -1$ is known, and s is chosen with $s \ge (q-1)/2$, then the balancing principle can be used, since from (2.1) we have the bound for the noise propagation (2.3). We shall not dwell into this. Instead we mention the following. For the balancing principle it is always an issue to choose the minimal regularization parameter α_0. We propose to run the balancing principle *after* the discrepancy principle, starting with $\alpha_0 := \alpha_D$, which for integer values q can be computed very fast by the ideas from [6]. The reason for this proposal is the following: In case of non-smooth solutions $x^\dagger$ that obey the smoothness assumption (1.3) with $p \le 2s+1-q$ we know from Theorem 2.2 that α_D provides the order optimal convergence rate. On the other hand, in case of smooth solutions $x^\dagger$ that obey the smoothness assumption (1.3) with $p > 2s+1-q$ we know from the discussion at the end of Section 2.5 that α_D is too small, and in such cases the balancing principle, in which the regularization parameter is chosen larger than α_D, will improve the convergence rate.

3 Concluding discussion

We conclude this study with discussing several topics, extending the previous analysis, and giving relation to other work in this direction. To this end we first recall the concept of general linear regularization in Hilbert space and the related concepts as these are general source conditions, and the framework of general noise assumptions, as this is used here. We will characterize the minimal error within this framework and we will then discuss in Section 3.1 how this best possible accuracy can be achieved by a certain *a priori* parameter choice, in particular, Remark 3.2 highlights that the maximal range (1.9) is recovered in the monomial framework.

One key aspect in this section is a discussion on how to generalize the Raus–Gfrerer rule for general linear regularization. Here we stick within the setup of power type source conditions and noise assumptions. However, we analyze a generalization of the RG-rule to the case of a general regularization scheme, say g_α, but the leading term of the RG-rule is still related to Tikhonov regularization, denoted by s_α, below. This gives some deeper insight into the nature of the Raus–Gfrerer rule, see also the final discussion at the end of this section.

We recall the following general setup from [8]. Below, we agree to call non-decreasing continuous functions $f\colon (0, \|A^*A\|] \to (0,\infty)$ with $\lim_{\lambda\to 0} f(\lambda) = 0$ *index functions*.

(i) Smoothness is given in terms of general source conditions as

$$x^\dagger \in H_\varphi := \left\{ x \in X \;\middle|\; x = \varphi(A^*A)v,\ \|v\| \le 1 \right\} \tag{3.1}$$

for an index function φ.

(ii) There is a function $\psi\colon (0, \|A^*A\|] \to \mathbb{R}^+$ such that the noise is bounded as

$$\left\| \psi(AA^*)\big(Ax^\dagger - y^\delta\big) \right\| \le \delta. \tag{3.2}$$

The minimal error, cf. (1.4), was characterized in [8, Theorem 4] as

$$\inf_{\mathcal{R}} \Delta(\delta, \mathcal{R}) \asymp \varphi\big(\Theta_{\psi\varphi}^{-1}(\delta)\big), \tag{3.3}$$

for the function $\Theta_{\psi\varphi}(\lambda) := \sqrt{\lambda}\psi(\lambda)\varphi(\lambda),\ 0 < \lambda \le \|A^*A\|$, provided that the latter is an increasing index function.

The authors in [8] considered linear regularization g_α given by piece-wise continuous functions, with residual functions $r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda)$ such that there are constants $\gamma_0, \gamma_* > 0$ such that for $0 < \lambda \le \|A^*A\|$, $0 < \alpha \le \bar{\alpha}$ we have

$$\lambda\, |g_\alpha(\lambda)| \le \gamma_0 \quad \text{and} \quad |g_\alpha(\lambda)| \le \frac{\gamma_*}{\alpha}.$$

Finally, it is assumed that there is a constant γ for which

$$|r_\alpha(\lambda)|\, \varphi(\lambda) \le \gamma\varphi(\alpha), \tag{3.4}$$

i.e., the regularization g_α has qualification φ with constant γ. The main results in [8] assert order optimality of regularization under a priori parameter choice, and also order optimality of the (version of the) discrepancy principle in [8, Theorem 5] under slightly enhanced qualification as usual, and the balancing principle in Theorem 6, ibid.

Here we aim at extending those results to linear regularization based on the generalization of (1.8) i.e., when

$$x_\alpha^\delta := g_\alpha\big(\left(A^*A\right)^{s+1}\big)\left(A^*A\right)^s A^* y^\delta, \quad 0 < \alpha \le \bar{\alpha}. \tag{3.5}$$

It is easy to see that the bias is given as

$$x^\dagger - x_\alpha = r_\alpha\big(\left(A^*A\right)^{s+1}\big)x^\dagger, \tag{3.6}$$

with element $x_\alpha := g_\alpha((A^*A)^{s+1})\,(A^*A)^{s+1}\,x^\dagger$. Thus we obtain the error decomposition as in (2.1).

This general type of regularization, but for the traditional noise assumption with $q = 0$ was already considered in [7]. Order optimal error bounds were obtained for the discrepancy and the balancing principles.

Within the present context it is convenient to assign to any index function φ the related $\varphi_s(\lambda) := \varphi(\lambda^{1/(s+1)}), 0 < \lambda \le \|A^*A\|^{(s+1)}$.

3.1 A priori parameter choice

The qualification assumption from (3.4), for the function φ_s yields that

$$\|x^\dagger - x_\alpha\| \le \gamma\varphi\big(\alpha^{1/(s+1)}\big), \quad 0 < \alpha \le \bar{\alpha}. \tag{3.7}$$

Similarly we can control the noise propagation

$$\begin{aligned} x_\alpha - x_\alpha^\delta &= g_\alpha\big(\left(A^*A\right)^{s+1}\big)\left(A^*A\right)^s A^*\left(Ax^\dagger - y^\delta\right) \\ &= g_\alpha\big(\left(A^*A\right)^{s+1}\big)\left(A^*A\right)^s A^*\psi\left(AA^*\right)^{-1}\psi\left(AA^*\right)\left(Ax^\dagger - y^\delta\right). \end{aligned}$$

In the light of (3.2) this allows to bound

$$\|x_\alpha - x_\alpha^\delta\| \le \delta\|g_\alpha\big(\left(A^*A\right)^{s+1}\big)\left(A^*A\right)^s A^*\psi\left(AA^*\right)^{-1}\|,$$

which in turn requires us to bound the function

$$\left|g_\alpha(\lambda^{s+1})\right| \frac{\lambda^{s+1/2}}{\psi(\lambda)} = \left|g_\alpha(\lambda^{s+1})\right| \frac{\sqrt{\lambda^{s+1}}}{\psi_s(\lambda^{s+1})\left(\lambda^{s+1}\right)^{-s/(2s+2)}}.$$

We claim that we can apply Lemma 2 in [8] with the function

$$\tilde{\psi}(\lambda) := \psi_s(\lambda)\lambda^{-s/(2s+2)}.$$

To this end we verify the assumptions that

$$\lambda\tilde{\psi}^2(\lambda) = \lambda\psi_s^2(\lambda)\lambda^{-s/(s+1)} = \lambda^{1/(s+1)}\psi^2(\lambda^{1/(s+1)})$$

is non-decreasing, and the same holds for

$$\frac{\lambda}{\tilde{\psi}^2(\lambda)} = \frac{\lambda\lambda^{s/(s+1)}}{\psi^2(\lambda^{1/(s+1)})} = \frac{(\lambda^{1/(s+1)})^{2s+1}}{\psi^2(\lambda^{1/(s+1)})},$$

which is non-decreasing if only the function $\lambda^{2s+1}/\psi^2(\lambda)$ was non-decreasing. Thus we have the following sufficient condition for the function ψ:

$$\text{The functions } \lambda \mapsto \lambda\psi^2(\lambda) \text{ and } \lambda \mapsto \frac{\lambda^{2s+1}}{\psi^2(\lambda)} \text{ are non-decreasing.} \tag{3.8}$$

If $s = 0$, then this is part of the assumption given in [8, Assumption A.5]. However, we see that if $s \geq -1/2$, then the benchmark (maximal) noise assumption is given by $\psi_{\max}(\lambda) := \lambda^{s+1/2}$. We thus can increase the range of applicability of regularization by increasing the value of s. Therefore, under (3.8) we can apply Lemma 2 in [8] to derive that

$$\left|g_\alpha(\lambda^{s+1})\right| \frac{\sqrt{\lambda^{s+1}}}{\tilde{\psi}(\lambda^{s+1})} \leq \frac{\max\{\gamma_0, \gamma_*\}}{\sqrt{\alpha^{1/(s+1)}}\psi(\alpha^{1/(s+1)})}, \tag{3.9}$$

cf. the above bounds (3.7) and (3.9) with (2.2) and (2.3), respectively. We summarize the above findings as

Theorem 3.1. *Suppose that g_α is a regularization which has qualification φ_s with constant γ, and that γ_0, γ_* are given as above. If $x^\dagger$ obeys* (3.1)*, the function ψ obeys* (3.2) *and* (3.8)*, and if x_α^δ is given as in* (3.5)*, then*

$$\|x^\dagger - x_\alpha^\delta\| \leq \gamma\varphi(\alpha^{1/(s+1)}) + \frac{\max\{\gamma_0, \gamma_*\}}{\sqrt{\alpha^{1/(s+1)}}\psi(\alpha^{1/(s+1)})}\delta.$$

The a priori parameter choice α_* from

$$\Theta_{\psi\varphi}\big(\alpha_*^{1/(s+1)}\big) = \delta \tag{3.10}$$

yields an order optimal regularization.

Remark 3.2. If the regularization g_α is given by Tikhonov regularization $g_\alpha(\lambda) = 1/(\alpha+\lambda)$, then φ_s has (maximal) qualification $\varphi_s(\lambda) = \lambda$ with constant 1. In terms of powers this translates to $0 < p/(2(s+1)) \le 1$, and hence gives $s \ge (p-2)/2$. The assumption (3.8) on the noise translates to $q \ge -1$ and $0 \le 2s + 1 - q$, and hence that $s \ge (q-1)/2$. Therefore, Theorem 3.1 fully recovers Proposition 2.1 as we require (1.9).

We notice, the following. If the regularization g_α has qualification p_0, i.e., we have (3.4) for the function $\varphi(t) = t^{p_0}$ then the range for s is enlarged from (1.9) to $2s + 2 \ge \max\left(\frac{p}{p_0}, q + 1\right)$, as we require that $0 < p/(2s+2) \le p_0$ in this case.

Remark 3.3. As already mentioned, we can increase the range of applicability of regularization by increasing the value of s. Notice that the parameter choice (3.10) with $s > 0$ will result in a regularization parameter which is smaller than the one for $s = 0$ for δ small enough, and hence the regularized problems tend to be more ill-conditioned along with increasing s.

Moreover, if a given regularization g_α has (finite and monomial) qualification $\varphi(\lambda) = \lambda^p$ for some $0 < p < \infty$, then it also has qualification φ_s. This can be deduced from

$$\frac{\varphi(\lambda)}{\varphi_s(\lambda)} = \frac{\lambda^p}{\lambda^{p/(s+1)}} = \lambda^{p\frac{s}{s+1}}, \quad \lambda > 0,$$

which is a non-decreasing function whenever $s \ge 0$. Therefore, by increasing s less qualification is required to cover higher smoothness.

3.2 A posteriori parameter choice

Along the same lines as in the previous subsection we obtain statements for the optimality of the discrepancy principle and the balancing principle.

We turn our attention to the Raus–Gfrerer rule, first given in Section 2.5. The regularization r_α in the leading operator $r_\alpha(G_s)^\mu$ in (2.20) need not be the one which is chosen to obtain x_α^δ, but this may always chosen from Tikhonov regularization. Indeed, to clearly distinguish, let $s_\alpha(\lambda) := \alpha/(\alpha + \lambda)$ by the residual function from Tikhonov regularization, and assume that we use some linear regularization, say g_α, to determine x_α^δ as in (3.5). The generalized Raus–Gfrerer rule is now given as solution, say α_{RG}, to

$$d_{\mathrm{RG}}(\alpha) := \left\| s_\alpha(G_s)^\mu (AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta)\right\| = C\delta. \tag{3.11}$$

If $r_\alpha(\lambda) := s_\alpha(\lambda)$, then this criterion coincides with the one from (2.20) used in Theorem 2.7. Again, for small data $\| (A^*A)^{q/2} (y - y^\delta)\| \le C\delta$, Remark 2.3 applies.

Remark 3.4. We stress that certain monotonicity properties of the functions g_α are required to determine the regularization parameter from (2.5) or (2.19), respectively. Otherwise the regularization parameter might be chosen from some finite geometric set $\alpha_j := \alpha_0 q^j$, $j = 1, \dots, M$, where $q > 1$ is some scaling factor, and α_0 is some minimal value of the regularization parameter.

The generalization of Theorem 2.7 is as follows.

Theorem 3.5. *Let x_α^δ be a regularized solution* (3.5) *for a regularization g_α with $|r_\alpha(\lambda)| \le \gamma_1$. Assume furthermore that* (3.4) *holds for the function $\varphi(\lambda) = \lambda^{p_0}$ for some $p_0 > 0$, and for some $\gamma \ge \gamma_1$. Let $s_\alpha(\lambda) := \alpha/(\lambda + \alpha)$ be from Tikhonov regularization. Let $\alpha := \alpha_{\mathrm{RG}}$ be chosen from equation* (3.11) *with $C > \gamma_1$ for $\mu := \frac{q+1}{2s+2}$. Under the conditions* (1.2) *and* (1.3)*, let s cover the (maximal) range $2s + 2 \ge \max\{p/p_0, q+1\}$. Then there is a constant $C(p, q, C, \gamma, \gamma_0, \gamma_1, \gamma_*)$ for which*

$$\|x^\dagger - x_{\alpha_{\mathrm{RG}}}^\delta\| \le C(p, q, C, \gamma, \gamma_0, \gamma_1, \gamma_*) E^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}.$$

The proof will follow the reasoning of the proof of Theorem 2.7, and we start with the following auxiliary observations.

Lemma 3.6. *Let g_α be a regularization for which equation* (3.4) *holds for the function $\varphi(\lambda) = \lambda^{p_0}$ for some $p_0 > 0$, and assume that $|r_\alpha(\lambda)| \le \gamma_1$. If $\gamma \ge \gamma_1$, then the estimate* (3.4) *holds with constant γ for each monomial function $\varphi(\lambda) = \lambda^\mu$ as long as $0 < \mu \le p_0$.*

Proof. Fix any $0 < \mu \le p_0$. Then the estimate (3.4) yields that

$$|r_\alpha(\lambda)|^{p_0/\mu} \lambda^{p_0} = |r_\alpha(\lambda)| \lambda^{p_0} |r_\alpha(\lambda)|^{p_0/\mu - 1} \le \gamma \alpha^{p_0} \gamma_1^{p_0/\mu - 1} \le \gamma^{p_0/\mu} \alpha^{p_0}.$$

The proof can be completed by taking the μ/p_0th power. □

Lemma 3.7. *Let g_α be any regularization, with residual function r_α, and let γ_1 be such that $\sup_{0<\lambda\le\|A^*A\|} |r_\alpha(\lambda)| \le \gamma_1$. Assume that for some $p_0 > 0$ estimate* (3.4) *holds for $\varphi(t) = t^{p_0}$. Then*

$$|r_\alpha(\lambda)|^{1/p_0} \le (\gamma^{1/p_0} + \gamma_1^{1/p_0}) \frac{\alpha}{\lambda + \alpha}, \quad 0 < \lambda \le \|H_1\|. \tag{3.12}$$

Consequently, we have for $0 < \mu \le 1$, and with $s_\alpha(\lambda) = \alpha/(\lambda + \alpha)$, that

$$\|r_\alpha(H_s)^{\mu/p_0} x\| \le (\gamma^{1/p_0} + \gamma_1^{1/p_0})^\mu \|s_\alpha(H_s)^\mu x\|, \quad x \in X. \tag{3.13}$$

Proof. If (3.4) holds for $\varphi(\lambda) = \lambda^{p_0}$ with constant γ, then

$$|r_\alpha(\lambda)|^{1/p_0} \lambda \le \gamma^{1/p_0} \alpha.$$

Thus we have that

$$(\lambda+\alpha)\,|r_\alpha(\lambda)|^{1/p_0} \le |r_\alpha(\lambda)|^{1/p_0}\,\lambda + \alpha\,|r_\alpha(\lambda)|^{1/p_0} \le \gamma^{1/p_0}\alpha + \gamma_1^{1/p_0}\alpha,$$

which gives (3.12). The inequality (3.12) implies that

$$\|r_\alpha(H_s)^{1/p_0}x\| \le (\gamma^{1/p_0} + \gamma_1^{1/p_0})\|s_\alpha(H_s)x\|, \quad x \in X.$$

Now we use the Löwner–Heinz inequality (operator monotonicity of the function $\lambda \mapsto \lambda^\mu$ for $0 < \mu \le 1$) to derive (3.13), which completes the proof. □

Proof of Theorem 3.5. First, we bound the noise propagation error as in (2.22), precisely, we have for $\alpha := \alpha_{\mathrm{RG}}$ that

$$\begin{aligned}
C\delta &= \left\|s_\alpha(G_s)^\mu \left(AA^*\right)^{q/2}(Ax_\alpha^\delta - y^\delta)\right\| = \left\|s_\alpha(G_s)^\mu r_\alpha(G_s)\left(AA^*\right)^{q/2} y^\delta\right\| \\
&\le \left\|s_\alpha(G_s)^\mu r_\alpha(G_s)\left(AA^*\right)^{q/2}(y-y^\delta)\right\| + \left\|s_\alpha(G_s)^\mu r_\alpha(G_s)\left(AA^*\right)^{q/2} y\right\| \\
&\le \left\|s_\alpha(G_s)^\mu r_\alpha(G_s)\right\|\delta + E\left\|s_\alpha(H_s)^\mu r_\alpha(H_s)\,(H_s)^{\frac{p+q+1}{2s+2}}\right\| \\
&\le \gamma_1\delta + E\left\|s_\alpha(H_s)^\mu r_\alpha(H_s)\,(H_s)^{\frac{p+q+1}{2s+2}}\right\|.
\end{aligned}$$

We distinguish two cases. If $\frac{p+q+1}{2s+2} \le p_0$, then (3.4) holds for $\varphi(\lambda) = \lambda^{\frac{p+q+1}{2s+2}}$, and we see, by virtue of Lemma 3.6, that

$$\left\|s_\alpha(H_s)^\mu r_\alpha(H_s)\,(H_s)^{\frac{p+q+1}{2s+2}}\right\| \le \gamma\left\|s_\alpha(H_s)^\mu\right\|\alpha^{\frac{p+q+1}{2s+2}} \le \gamma\alpha^{\frac{p+q+1}{2s+2}}.$$

Otherwise, if $\frac{p+q+1}{2s+2} > p_0$, then we argue that

$$\begin{aligned}
\left\|s_\alpha(H_s)^\mu r_\alpha(H_s)\,(H_s)^{\frac{p+q+1}{2s+2}}\right\| &= \left\|s_\alpha(H_s)^\mu\,(H_s)^{\frac{p+q+1}{2s+2}-p_0} r_\alpha(H_s)H_s^{p_0}\right\| \\
&\le \gamma\alpha^{p_0}\left\|s_\alpha(H_s)^\mu\,(H_s)^{\frac{p+q+1}{2s+2}-p_0}\right\|.
\end{aligned}$$

Now we observe that $\frac{p+q+1}{2s+2} - p_0 \le \mu$, since

$$\frac{p+q+1}{2s+2} - p_0 = \frac{p}{2s+2} + \frac{q+1}{2s+2} - p_0 \le p_0 + \mu - p_0 = \mu,$$

such that also in this case we have $\|s_\alpha(H_s)^\mu r_\alpha(H_s)\,(H_s)^{\frac{p+q+1}{2s+2}}\| \le \gamma\alpha^{\frac{p+q+1}{2s+2}}$. We thus arrive at a lower bound for α_{RG} as

$$\alpha_{\mathrm{RG}} \ge \left(\frac{C-\gamma_1}{\gamma E}\delta\right)^{\frac{2s+2}{p+q+1}}. \tag{3.14}$$

The bound for the noise propagation error is then given from (3.9) by

$$\left\| x_{\alpha_{\mathrm{RG}}}^{\delta} - x_{\alpha_{\mathrm{RG}}} \right\| \le \max\left\{\gamma_0, \gamma_*\right\} \left(\frac{\gamma E}{C - \gamma_1} \right)^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}. \tag{3.15}$$

We turn to bounding the regularization error. We use Lemma 3.7 with the element $x := r_\alpha(H_s) \left(A^* A\right)^{\frac{q+1}{2}} x^\dagger$ and find that

$$\begin{aligned} &\left\| r_\alpha(H_s)^{\mu/p_0} r_\alpha(H_s) \left(A^* A\right)^{\frac{q+1}{2}} x^\dagger \right\| \\ &\qquad \le \left(\gamma^{1/p_0} + \gamma_1^{1/p_0}\right)^{\mu} \left\| s_\alpha(H_s)^{\mu} r_\alpha(H_s) \left(A^* A\right)^{\frac{q+1}{2}} x^\dagger \right\| \\ &\qquad \le \left(\gamma^{1/p_0} + \gamma_1^{1/p_0}\right)^{\mu} (C + \gamma_1)\delta, \end{aligned}$$

which is obtained similar to the proof of (2.24). We shall apply the interpolation type argument. To this end we note that $p/(p_0(2s+2)) \le 1$, and we apply the reasoning of (2.25)–(2.27) from the proof of Theorem 2.7 to

$$z := r_\alpha^{1 - \frac{p}{p_0(2s+2)}}(H_s) v, \text{ and operator } B := r_\alpha^{\frac{1}{p_0(2s+2)}}(H_s) \left(A^* A\right)^{1/2}.$$

The interpolation arguments give the following bound for the bias:

$$\left\| x^\dagger - x_\alpha \right\| \le \left(\left(\gamma^{1/p_0} + \gamma_1^{1/p_0}\right)^{\mu} (C + \gamma_1)\delta \right)^{\frac{p}{p+q+1}} (\gamma_1 E)^{\frac{q+1}{p+q+1}}.$$

This results in a bound for the overall error under the generalized Raus–Gfrerer rule given by

$$\left\| x^\dagger - x_{\alpha_{\mathrm{RG}}}^{\delta} \right\| \le C E^{\frac{q+1}{p+q+1}} \delta^{\frac{p}{p+q+1}}$$

with constant

$$\begin{aligned} C := \Bigg(\left(\left(\gamma^{1/p_0} + \gamma_1^{1/p_0}\right)^{\mu} (C + \gamma_1) \right)^{\frac{p}{p+q+1}} \gamma_1^{\frac{q+1}{p+q+1}} \\ + \max\left\{\gamma_0, \gamma_*\right\} \left(\frac{\gamma}{C - \gamma_1} \right)^{\frac{q+1}{p+q+1}} \Bigg), \end{aligned}$$

which completes the proof. □

Remark 3.8. Notice that for $r_\alpha = s_\alpha$, i.e., when Tikhonov regularization is used, then we have that $\gamma = \gamma_1 = \gamma_0$, $\gamma_* \le 1$. Since then Lemma 3.7 holds with constant 1 instead of $(\gamma^{1/p_0} + \gamma_1^{1/p_0})$, we recover the bound from Theorem 2.7.

We summarize the discussion in this section. We extended results from Section 2 to more general linear regularization schemes, which are given by a function g_α and determine the reconstructions as in (3.5). We considered more general noise assumptions as expressed in (3.2) with a function ψ for which (3.8) hold, and general smoothness as given by (3.1). The benchmark order of reconstruction is given in (3.3).

(I) Theorem 3.1 provides order optimal reconstruction for an appropriate *a priori* parameter choice for all smoothness situations φ of the smoothness assumption (3.1) for which $\varphi_s(\lambda) := \varphi(\lambda^{1/(s+1)})$ is a qualification of $g_\alpha(\lambda)$.

(II) We mentioned that an appropriate modification of the discrepancy principle (2.5) with $C > 1$ for choosing α, following the proof of Theorem 2.4 within the more general setup of this section, also yields order optimal reconstruction, however, for functions φ of the smoothness assumption (3.1) for which $\Theta_{\psi\varphi}(\lambda^{1/(s+1)})$ is a qualification of $g_\alpha(\lambda)$. For the proof of this result we additionally require that the function $\lambda \mapsto \varphi^2\big((\Theta_{\psi\varphi})^{-1}(\lambda)\big)$ is concave. Please note that the set of index functions φ that obey the latter assumption is smaller compared with the set of index functions that work in Item (I). This reflects that the discrepancy principle saturates too early. In the special case of Section 2, the latter qualification assumption is satisfied for the p-range $0 < p \le 2s + 1 - q$.

(III) We also considered the RG-rule (3.11) with $C > 1$. If g_α is from Tikhonov regularization, hence $r_\alpha = s_\alpha$ in this case, then the rule (3.11) coincides with (2.20) for $s = q = 0$. Otherwise this is not the case, and the rule (3.11) is one possible modification/generalization of (2.20) in case that the general regularization which has qualification at least $\varphi(\lambda) = \lambda^{p_0}$ for some $p_0 > 0$. However, we used (a power of) Tikhonov regularization in front of the residual in order to prevent early saturation, see (3.11). Letting $\mu = (q+1)/(2s+2)$ for choosing α, and following the ideas of the proof of Theorem 2.7, we obtain that the early saturation of the discrepancy principle can be prevented and that order optimal error bounds can be guaranteed for all smoothness and noise situations, as in case of *a priori* parameter choice. This value of μ only uses the known noise assumption q and the chosen power s, it is thus feasible.

One may also consider some rule different from (3.11), and replace s_α by r_α, i.e., the parameter α_{RG} is chosen from the nonlinear equation

$$\left\| r_\alpha(G_s)^\mu (AA^*)^{q/2}(Ax_\alpha^\delta - y^\delta) \right\| = C\delta \quad \text{with } \mu = (q+1)/[2p_0(s+1)].$$

In the special case $s = 0$ and $q = 0$, this rule coincides with the classical RG rule. In analogy to the proof of Theorem 2.7 it can be shown that also for this a posteriori rule the order optimality result of Theorem 3.5 holds true for the maximal range

$$2s + 2 \ge \max\{p/p_0, q+1\},$$

and we skip the details here. In both case we restricted to the setup of monomial smoothness and noise assumptions.

Bibliography

[1] D. D. Ang, R. Gorenflo, V. K. Le and D. D. Trong, *Moment Theory and Some Inverse Problems in Potential Theory and Heat Conduction*, Lecture Notes in Mathematics 1792, Springer-Verlag, Berlin, 2002.

[2] G. Blanchard and P. Mathé, Conjugate gradient regularization under general smoothness and noise assumptions, *J. Inverse Ill-Posed Probl.* **18** (2010), no. 6, 701–726.

[3] H. Egger, Regularization of inverse problems with large noise, *J. Phys. Conf. Ser.* **124** (2008), 012022.

[4] P. P. B. Eggermont, V. N. Lariccia and M. Z. Nashed, On weakly bounded noise in ill-posed problems, *Inverse Problems* **25** (2009), no. 11, 115018.

[5] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[6] S. Lu, S. V. Pereverzev, Y. Shao and U. Tautenhahn, On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales, *J. Integral Equations Appl.* **22** (2010), no. 3, 483–517.

[7] P. Mathé and U. Tautenhahn, Interpolation in variable Hilbert scales with application to inverse problems, *Inverse Problems* **22** (2006), 2271–2297.

[8] P. Mathé and U. Tautenhahn, Regularization under general noise assumptions, *Inverse Problems* **27** (2011), no. 3, 035016.

[9] V. A. Morozov, Regularization under large noise, *Comput. Math. Math. Phys.* **36** (1996), 1175–1181.

[10] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Wiley, New York, 1977.

[11] G. M. Vainikko and A. Y. Veretennikov, *Iteration Procedures in Ill-Posed Problems*, Nauka, Moscow, 1986 (in Russian).

Received April 20, 2011; revised May 27, 2011.

Author information

Peter Mathé, Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstrasse 39, 10117 Berlin, Germany.
E-mail: mathe@wias-berlin.de

Ulrich Tautenhahn, Department of Mathematics, University of Applied Sciences
Zittau/Görlitz, P. O. Box 1454, 02754 Zittau, Germany.
E-mail: u.tautenhahn@hs-zigr.de